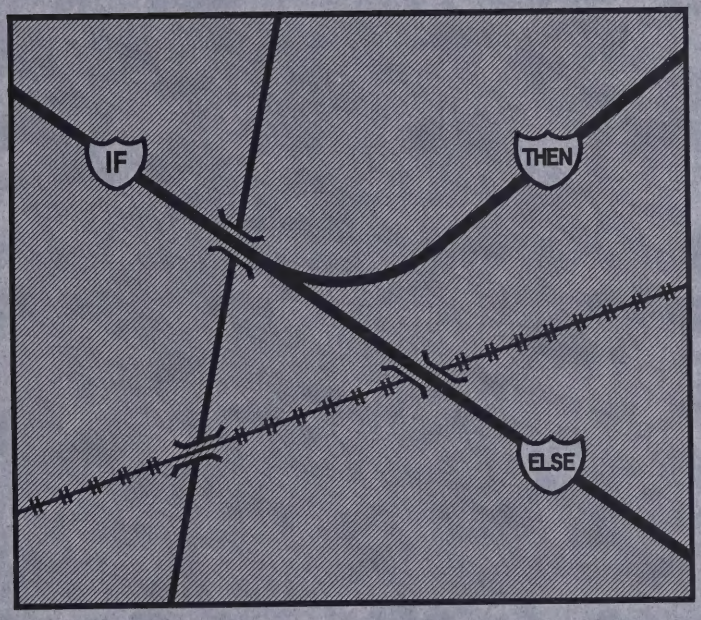
97-001

HIGHWAY QUALITY ASSURANCE:

THE NEED FOR SCIENCE-AND-KNOWLEDGE-BASED DECISION MAKING



RICHARD M. WEED
NEW JERSEY DEPARTMENT OF TRANSPORTATION

MARCH 1997

NJ
TE
175
W44
1997
c. 1

1. Report Number FHWA/NJ - 97 - 001 - 7490	2. Government Accession Number	3. Recipient Catalog Number
4. Report: Highway Quality Assurance: The Need for Science-and-Knowledge-Based Decision Making		5. Report Date March 1997
7. Author: Richard M. Weed (See Acknowledgment in Item 15)		6. Performing Agency Code N/A
9. Performing Agency Name and Address New Jersey Department of Transportation 1035 Parkway Avenue Trenton, NJ 08625		8. Performing Agency Report Number NJ - 97 - 001 - 7490
12. Sponsoring Agency Name and Address Federal Highway Administration U. S. Department of Transportation Washington, DC		10. Work Unit Number (TRAIS) N/A
15. Supplementary Notes (ACKNOWLEDGMENT) Several individuals contributed significantly to this report -- writing some portions, editing others, and reviewing all sections of the finished product. They are Kenneth C. Afferton, former Assistant Commissioner, New Jersey Department of Transportation; Dr. James L. Burati, Professor of Civil Engineering, Clemson University; Dr. Michael I. Darter, Vice President, ERES Consultants, Inc.; Charles S. Hughes, Consultant; and Dr. Olga J. Pendleton, Pen-Hock Statistical Consultants. The author gratefully acknowledges their valuable assistance as both mentors and colleagues with a shared vision -- to improve the practice of highway quality assurance.		11. Contract or Grant Number NJ SPR Study 7490
		13. Type of Report/Period Covered Final
		14. Sponsoring Agency Code

16. Abstract

Despite many advances in technology over the past two decades, and a recently renewed commitment to quality in the form of the National Quality Initiative (NQI), there remain many quality assurance procedures in use today that reflect neither the proper application of technology nor the commitment to excellence proclaimed by the NQI. It is argued that critical decision-making processes in the highway field -- processes that affect the acceptance or rejection of work performed by the construction industry, or the level of compensation for such work -- must be based on sound engineering and mathematical principles. Several examples are presented to illustrate that the best way to assure that good decisions will be made is to insist on the use of scientifically sound procedures. Recommendations are also made to improve the process by which voluntary consensus standards are produced. The consensus process is well suited to resolve many issues, but it may not be conducive to good science. Scientific procedures must stand or fall on their technical merits. Careful scrutiny by knowledgeable individuals must be invited and welcomed, and valid criticism acted upon. Only in this manner can the process be made productive and credible.

17. Key Words Quality assurance, statistical quality assurance, statistical specifications, adjusted pay schedules, operating characteristic curves, statistical decision making, statistical misapplications.	18. Distribution Statement		
19. Security Class (This Report) Unclassified	20. Security Class (This Page) Unclassified	21. Total Pages 30	22. Price

The contents of this report reflect the views of the author who is responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily represent the official views or policies of the New Jersey Department of Transportation or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

CONTENTS

INTRODUCTION 1
COSTS OF POOR PRACTICES 2
STATISTICAL TOOLS 3
UNDULY LENIENT ACCEPTANCE PROCEDURE 9
UNDULY SEVERE ACCEPTANCE PROCEDURE 11
WEAKNESS OF SINGLE-TEST PROCEDURE 13
CASE IN WHICH UNIFORMITY MAY NOT BE BENEFICIAL 18
WASTEFUL METHODS 21
IMPROVING THE VOLUNTARY CONSENSUS PROCESS 24
SUMMARY AND CONCLUSIONS 27
REFERENCES 29

FIGURES

FIGURE 1 Conceptual representation of accuracy, precision, and bias 4
FIGURE 2 Conventional OC curve 7
FIGURE 3 Typical OC curve for acceptance procedure with an adjusted pay schedule 8
FIGURE 4 Illustration of weakness of decision procedure based on a single test 17
FIGURE 5 Illustration of case in which uniformity may not be beneficial 19

TABLES

TABLE 1 Pay schedule for unduly lenient acceptance procedure 10
TABLE 2 Operating characteristics of unduly lenient acceptance procedure 10
TABLE 3 Operating characteristics of unduly severe acceptance procedure 12
TABLE 4 Range coefficients for use with single-test decision procedure 14
TABLE 5 Operating characteristics of single-test decision procedure 16
TABLE 6 Comparative performance of four acceptance plans 23

The Quality Mission...

To foster a greater sense of vision and direction in the application of highway quality assurance, to advocate the correct use of scientific and mathematical principles in the development of practices and procedures consistent with the same high standards of excellence being sought in the finished product, to uphold the fair and ethical treatment of all contractual participants, to support people and organizations working to achieve these goals, and to develop professional relationships founded on competence and integrity.

INTRODUCTION

The keynote paper by Hunter and Pendleton (1) at the 1992 TRB session on "Managing Quality: A National Perspective" outlined in a conceptual way the importance of statistical science in transportation. The authors stressed that "quality transportation *technology* must accompany quality transportation *management,*" and further noted that transportation leaders must insist upon the use of valid statistical procedures if transportation technology is to keep pace with modern science.

However, despite many advances in technology over the past two decades, and a recently renewed commitment to quality in the form of the National Quality Initiative (NQI) (2), there remain many quality assurance procedures in use today that reflect neither the proper application of technology nor the commitment to excellence proclaimed by the NQI. Critical decision-making processes in the highway quality assurance field -- processes that affect the acceptance or rejection of work performed by the construction industry, or the level of compensation for such work -- have not always been based on the proper application of sound engineering and mathematical principles. In some cases, well intentioned but misguided attempts to simplify standard statistical procedures have seriously compromised their technical integrity. In other cases, basically sound procedures have been used inappropriately. Often, no checks have been conducted to confirm that the applications are effective. Several examples are presented to illustrate these shortcomings.

Of even greater concern is the harm that is done when inadequate procedures appear in national guide specifications that serve as models for practitioners who may not have a

complete understanding of the underlying theory. Much of what has been published on highway quality assurance in these documents is lacking in technical rigor and often has not been subjected to any formal analysis to demonstrate its adequacy. As a result, many procedures in use today are inefficient, some are ineffective, and a few may even be counterproductive, providing greater reward for lesser quality. This should be cause for alarm and, if it is to be rectified, changes must be made in the process by which standards and guide specifications are produced. Suggestions are offered to address this problem.

COSTS OF POOR PRACTICES

The use of correct practices in highway quality assurance is of more than mere academic interest. Poorly developed programs do little but breed disrespect, distrust, a generally adversarial atmosphere, and a determination on the part of the construction industry to oppose the programs. In a discipline dedicated to excellence, it is totally inappropriate to tolerate either specifications or consensus standards that are far from excellent. If demands for excellence are to be placed upon the construction industry, then transportation agencies must be no less diligent in placing the same demands upon themselves in the development of the specifications and standards that govern this work.

But the costs of poor practices go beyond the loss of integrity and credibility. The replacement value of the nation's roads and bridges has been estimated to be between $1 and $3 trillion (3). The use of sound quality assurance practices is one of the keys to protecting this investment. No detailed analysis is required to see that an ineffectiveness of only a few

percent could result in the loss of millions, if not billions, of dollars in terms of lost performance.

If the transportation profession is to retain credibility in the field of highway quality assurance, it must be continually critical of its own performance, must be prepared to acknowledge mistakes, and must act quickly to correct any discovered deficiencies. Larson and Albright said it well in a recent paper on transportation research (4, page 164): "Competence and integrity must invite challenge -- and so be strengthened."

STATISTICAL TOOLS

To demonstrate that not all statistical procedures perform equally well, statisticians have developed a variety of quantitative tools, such as the concepts of accuracy, precision, and bias.

A procedure is said to be *accurate* if the mean of a distribution of measurements tends to coincide with the true mean of the population. A procedure is said to be *precise* if repeat tests or measurements on identical samples tend to reproduce the same value. *Bias* is a measure of inaccuracy, that is, the degree to which the mean of a distribution of measurements tends to be displaced from the true population mean.

Perhaps the best visual impression of these terms can be obtained by imagining a marksman shooting at a target (5), as shown in Figure 1. The first case, in which the measurements are both accurate and precise, represents the conditions that are desired, enabling the transportation agency to make fair and appropriate acceptance decisions.

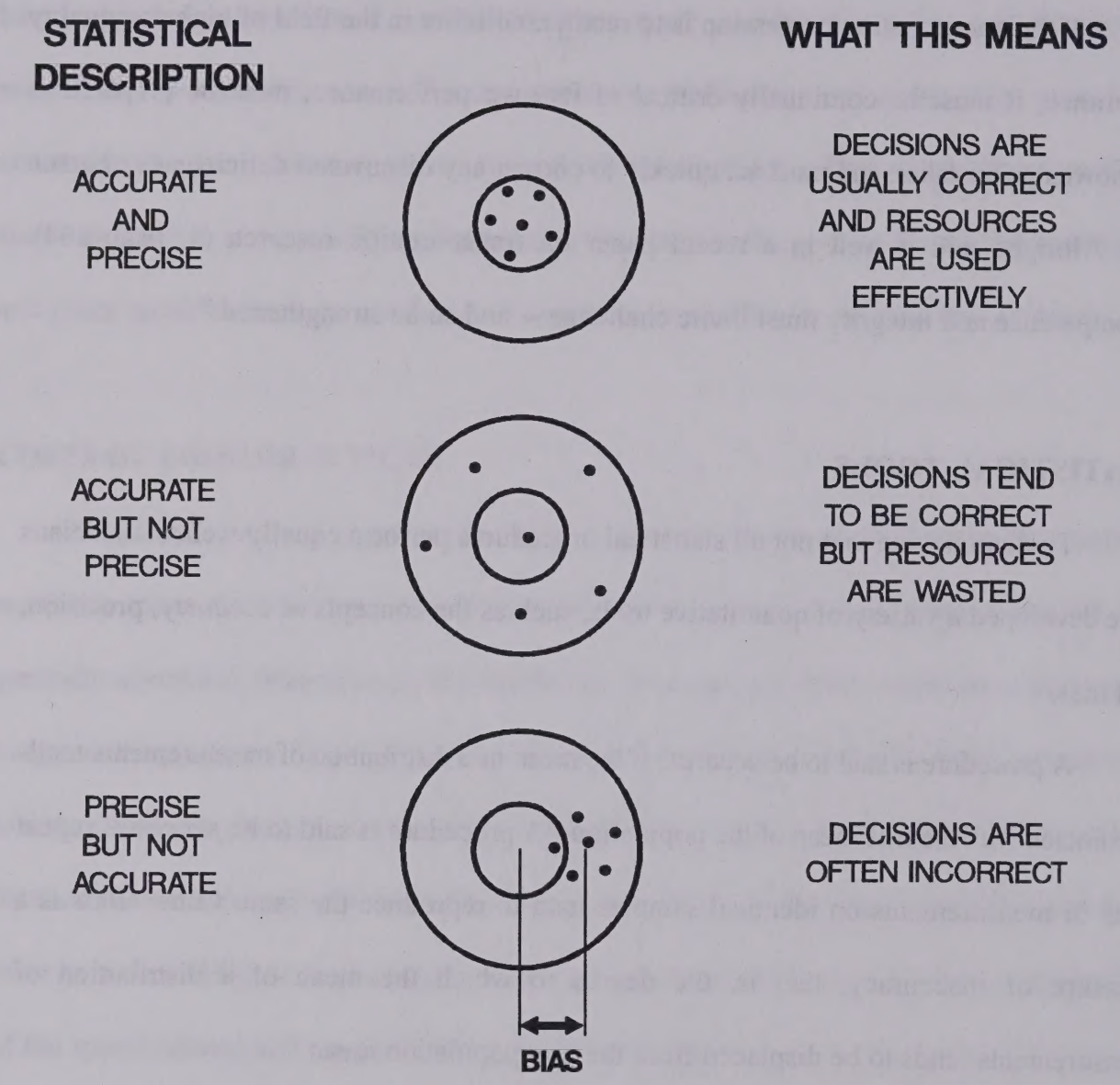

STATISTICAL DESCRIPTION
WHAT THIS MEANS
ACCURATE AND PRECISE
DECISIONS ARE USUALLY CORRECT AND RESOURCES ARE USED EFFECTIVELY
ACCURATE BUT NOT PRECISE
DECISIONS TEND TO BE CORRECT BUT RESOURCES ARE WASTED
PRECISE BUT NOT ACCURATE
DECISIONS ARE OFTEN INCORRECT
BIAS

Figure 1. Conceptual representation of accuracy, precision, and bias.

The second example in Figure 1 shows a process that is accurate but not precise. This is less desirable, of course, but does not automatically rule out the use of such procedures. If the less precise procedure happens to be simpler, faster, or less expensive, and the consequence of making an incorrect decision is not particularly great, then the less precise procedure may be adequate. Or, if the decision is of critical importance that would warrant greater precision, this can still be accomplished with the less precise procedure by using a larger sample size which improves the precision of the overall process.

The third example in Figure 1 is more problematical. Bias is a particularly insidious problem for two reasons: it can greatly distort the results that are obtained, and there usually is no obvious warning that it is present because the true population parameters are almost never known. Although there are rare situations in which statisticians have been willing to tolerate a small amount of bias in exchange for a substantial improvement in precision, such a compromise should be approached with extreme caution. As a minimal defense against bias, only well established and proven statistical procedures should be used.

Another extremely useful tool for the analysis of statistical acceptance procedures is the operating characteristic (OC) curve. It is only through the study of such curves that the risks to both the agency and the contractor can be quantified and controlled at suitably low levels. This enables the agency to develop fair and effective specifications and may aid the contractor in determining the appropriate bidding and production strategies.

Even though the acceptance procedure or pay schedule spells out precisely the decision to be made for any level of measured quality, there is always some degree of

uncertainty in the quality measurement itself. This occurs because only a small fraction of the lot is sampled and tested. The OC curve, if constructed properly, is capable of accounting for this uncertainty.

A conventional OC curve is shown in Figure 2. Probability of acceptance is indicated on the vertical axis for the range of quality levels indicated schematically on the horizontal axis. The contractor's risk of having good (acceptable quality level, AQL) material rejected and the agency's risk of accepting poor (rejectable quality level, RQL) material are both illustrated in this figure.

Figure 3 presents an OC curve constructed for a statistical specification with an adjusted pay schedule. Quality levels are indicated on the horizontal axis in the usual manner but, instead of probability of acceptance, the vertical axis gives the expected pay factor.

Although the risks have a slightly different interpretation when associated with the expected payment curve in Figure 3, essentially the same type of information is provided. In this particular example, AQL work receives an expected pay factor of 100 percent, as desired, while truly superior work receives an incentive pay factor slightly above 100 percent. At the other extreme, RQL work corresponds to an expected pay factor of 70 percent. Presumably, the transportation agency has determined that this will cause sufficient money to be withheld to cover the anticipated cost of future repairs. For still lower levels of quality, this curve levels off at the minimum pay factor of 50 percent.

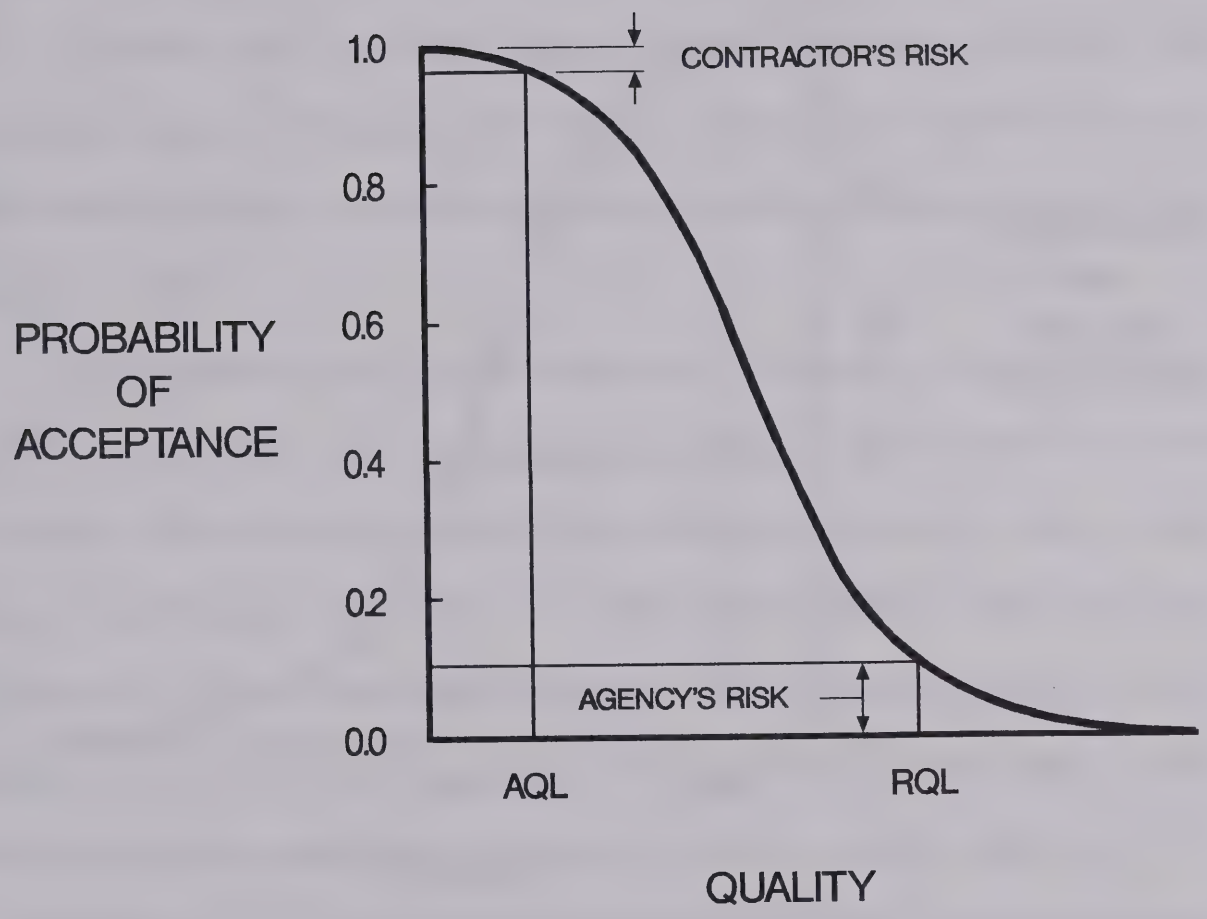


Figure 2. Conventional OC curve.

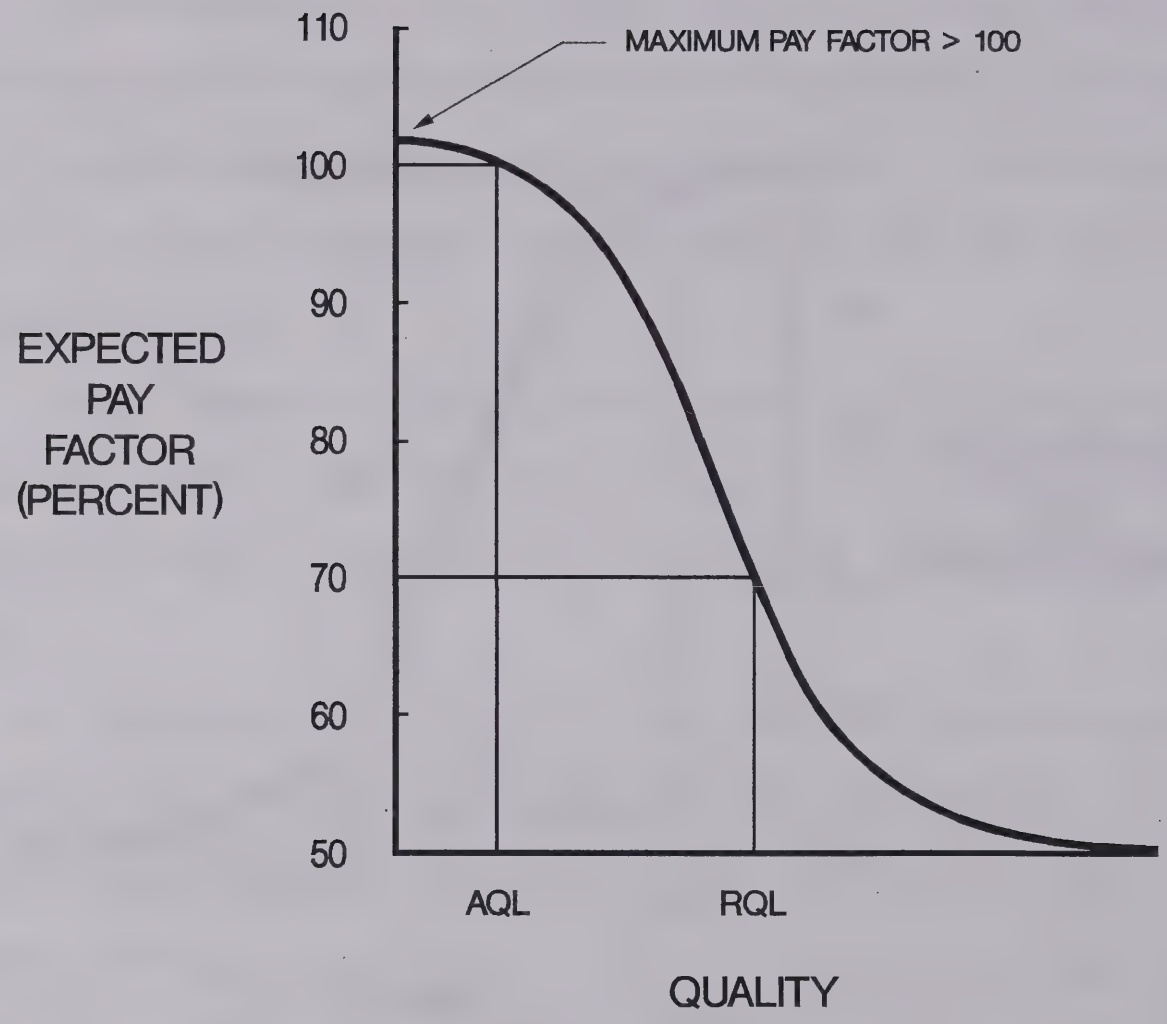


Figure 3. Typical OC curve for statistical acceptance procedure with an adjusted pay schedule.

These tools -- the OC curve and the concepts of accuracy, precision, and bias -- will be seen to be instrumental in exposing the weaknesses of the questionable procedures that follow.

UNDULY LENIENT ACCEPTANCE PROCEDURE

The most common oversight in the development of statistical specifications is the failure to construct the OC curves to verify that the acceptance plans will perform as desired. The following example is taken from an acceptance procedure that appeared for a time in a national standard.

This particular generic acceptance procedure was based on percent within limits (PWL), the percentage of the lot estimated to fall within specification limits. It used the standard deviation method to estimate the PWL, and it included an incentive clause to reward superior performance. In the text of the original standard, the AQL was defined as 95 percent within limits. A stepped pay schedule provided for incentive pay factors up to 105 percent for greater quality and proportionally lower pay factors for lesser quality. Table 1 lists the pay factors assigned by the acceptance procedure for a typical sample size of N = 5.

When the OC curves for this procedure were checked, however, it was found to perform much differently than its developers expected. The OC curve for the pay schedule in Table 1 is given in tabular form in Table 2.

Table 1. Pay schedule for unduly lenient acceptance procedure.

MINIMUM PERCENT WITHIN LIMITS	PAY FACTOR (PERCENT)	MINIMUM PERCENT WITHIN LIMITS	PAY FACTOR (PERCENT)
100	105	59	89
92	104	57	88
87	103	56	87
83	102	55	86
80	101	53	85
78	100	52	84
75	99	51	83
73	98	50	82
71	97	48	81
69	96	47	80
68	95	46	79
66	94	45	78
65	93	43	77
63	92	42	76
62	91	41	75
60	90	< 41	Reject[a]

[a] Pay factor assumed to be zero for analysis purposes

Table 2. Operating characteristics of unduly lenient acceptance procedure.

PERCENT WITHIN LIMITS	AVERAGE PAY FACTOR (PERCENT)
100	105.0
95 (AQL)	103.7
90	102.2
85	100.5
80	97.8
75	94.3
70	89.9

It can be seen from Table 2 that the contractor who performs consistently at the AQL of 95 percent within limits will receive an average pay factor of nearly 104 percent. Furthermore, the contractor who furnishes work that is approximately 85 percent within limits, substantially below the level that was defined as acceptable, will still receive an average pay factor of about 100 percent. It is doubtful that any agency would regard a known and predictable overpayment such as this to be appropriate and, accordingly, the standard was revised.

The result at the AQL may be regarded as a form of bias, as illustrated in the third diagram in Figure 1. As is often the case with bias of this type, it became apparent only after the OC curve analysis was performed.

UNDULY SEVERE ACCEPTANCE PROCEDURE

A second example, which had also been proposed for use in a national standard, illustrates how unduly severe a pay schedule can be if it is not designed properly. Like the previous example, the statistical measure of quality was the PWL except that, in this case, the AQL was defined as 90 percent within limits. Although the pay adjustment was initially written in terms of percent reduction, it is presented in Equation 1 in terms of percent pay factor so that it can be analyzed with the aid of recently developed software (6).

$$PF = 10 + PWL \quad (\text{Subject to } PF \leq 100) \tag{1}$$

where

PF = pay factor (percent) and

PWL = percent within limits computed from test values.

Again using a typical sample size of N = 5, the operating characteristic curve for this pay schedule is presented in Table 3. Here it can be seen that the problem is just the opposite of the one discovered in the previous example. The values in Table 3 show that a contractor who performs consistently at the AQL will receive an average pay reduction of nearly 5 percent. To emphasize the impact this would have on the construction industry, a contractor responsible for $10 million worth of work under this specification over the course of a construction season would be penalized approximately $500,000 for successfully providing the level of quality that was explicitly defined as acceptable in the contract documents.

Table 3. Operating characteristics of unduly severe acceptance procedure.

PERCENT WITHIN LIMITS	AVERAGE PAY FACTOR (PERCENT)
100	100.0
90 (AQL)	95.2
80	87.5
70	79.1
60	69.5
50	59.7

This case, also, is an example of bias that would have gone undetected had the OC curve analysis not been performed. To correct this problem so that an average pay factor of 100 percent will be received at the AQL, an incentive provision (pay factor greater than 100 percent for truly superior quality) must be made a part of the acceptance procedure (6, Example 3.4).

WEAKNESS OF SINGLE-TEST PROCEDURE

One of the most important factors influencing the ability of statistical procedures to produce correct decisions is the amount of data available. The more critical the decision, the more important it is to have a sufficiently large sample to provide reasonable assurance that a correct decision will be made.

This fact can be illustrated with a procedure that is currently being considered for use in a national standard. Its purpose is to determine whether or not a set of test results obtained by a contractor is consistent with a single test performed by a state agency. Based on the range (R) of the contractor's data set, the test requirement is given by Equation 2, and the appropriate range coefficient (C) is obtained from Table 4.

$$\bar{X} - C^*R \leq X \leq \bar{X} + C^*R \qquad (2)$$

where

$\bar{X}$ = average of several quality control tests performed by the contractor,

R = range of the quality control tests,

C = coefficient by which the range is multiplied, and

X = single quality assurance test performed by the transportation agency.

Table 4. Range coefficients for use with single-test decision procedure.

NUMBER OF CONTRACTOR'S TESTS (N)	RANGE COEFFICIENT (C)
5	1.61
6	1.33
7	1.17
8	1.05
9	0.97
10	0.91

To evaluate a decision procedure such as this, there are three key questions that must be answered:

1. What is the probability of falsely detecting a difference between the two data sets when there truly is no difference?

2. What is the probability of correctly detecting a difference between the two data sets when such a difference truly exists? For example, when considering the expected performance of concrete pavement or structures, a

deficiency in compressive strength of 6890 kPa (1000 psi) might be associated with a significant loss in service life or performance. Therefore, it would be very beneficial to know the likelihood of correctly detecting a difference of this magnitude.

3. Similarly, it will be useful to know how large the true difference between two populations must be for there to be a reasonably high probability that it will be detected. For example, again using concrete compressive strength, what must the true difference in population means be for there to be a 0.90 probability that it will be detected?

To answer these questions, an OC curve analysis must be performed, the results of which are shown in Table 5 for the case in which there are N = 5 contractor's tests. The true difference between the means of the two populations is given in standard deviation units so that these results can be generalized to a wide variety of applications.

The answer to the first question can be read from the first row in Table 5. The risk of 0.02 of falsely detecting a difference when none exists would probably be tolerable for most applications.

The answers to the second and third questions reveal a serious shortcoming, however. For example, depending upon the length of time over which the data are gathered, the standard deviation associated with concrete compressive strength may be as large

Table 5. Operating characteristics of single-test decision procedure.

DIFFERENCE IN MEANS (SD UNITS)	PROBABILITY OF DETECTING DIFFERENCE
0.0	0.02
1.0	0.05
2.0	0.17
3.0	0.35
4.0	0.57
5.0	0.76
6.0	0.90

as 3445 kPa (500 psi). To answer the second question, a difference of 6890 kPa (1000 psi) would be equivalent to 2.0 standard deviation units and, using the results in Table 5, the probability of detecting a difference this large when it truly exists is only about 0.17. To answer the third question, Table 5 indicates that a true difference of 6.0 standard deviation units is necessary for there to be a 0.90 probability of detection. Therefore, the actual difference in concrete compressive strength must be 6.0 x 3445 = 20,670 kPa (3000 psi) in order for there to be a reasonably high probability of detection.

Figure 4 illustrates how this procedure readily leads to the conclusion that there is no difference between the two data sets, even when the two populations are completely separated. Obviously, there are not many engineering applications for which such a weak procedure could be considered practically useful. In fact, it could provide a false sense of security by continually finding no significant difference between the data sets when the true differences are actually quite large. In many cases, this could allow seriously defective work to be accepted without the agency being aware of it.

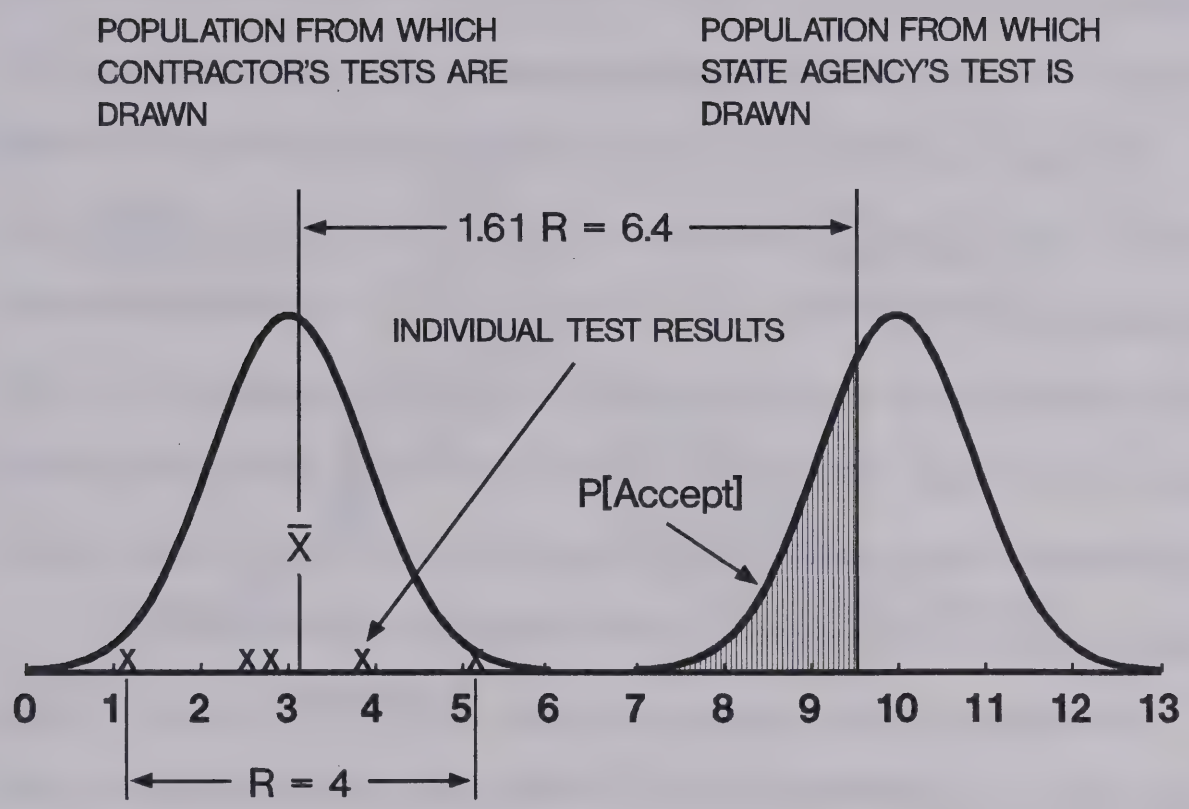

POPULATION FROM WHICH CONTRACTOR'S TESTS ARE DRAWN
POPULATION FROM WHICH STATE AGENCY'S TEST IS DRAWN
1.61 R = 6.4
INDIVIDUAL TEST RESULTS
P[Accept]
$\bar{X}$
0 1 2 3 4 5 6 7 8 9 10 11 12 13
R = 4

Figure 4. Illustration of weakness of decision procedure based on a single test.

CASE IN WHICH UNIFORMITY MAY NOT BE BENEFICIAL

Another procedure that is under consideration for inclusion in a national standard would provide additional reward (incentive pay provision) to the contractor who achieves greater uniformity of production. While there certainly are construction situations for which uniformity is beneficial, there are also situations in which it may actually be detrimental.

The Taguchi Loss Function (7) is a concept whereby product is not regarded as simply "good" or "bad" but, rather, is evaluated as a function of its *distribution* with respect to specification limits. Material falling just within specification limits (or just on the acceptable side of a single limit) is believed to provide less functionality than material that is within limits by a wider margin. (Although arrived at intuitively, the use by the highway profession of pay adjustments based on percent defective or percent within limits as the quality measure predates Taguchi by nearly two decades on this extremely perceptive insight.)

For example, highway specifications are often based on the percentage of material falling below a single lower limit when it is known that performance increases as some construction variable increases (e.g., pavement fatigue life increases with increasing strength or thickness). At generally satisfactory quality levels, it is highly unlikely that greater uniformity yields greater benefits in terms of performance.

To see why this is so, it is necessary to compare two pavement distributions of *identical quality in terms of the statistical measure prescribed for the proposed application.* Figure 5 shows two normal distributions that are accurately drawn, each of which has exactly 10 percent defective (90 percent within limits), a commonly used value

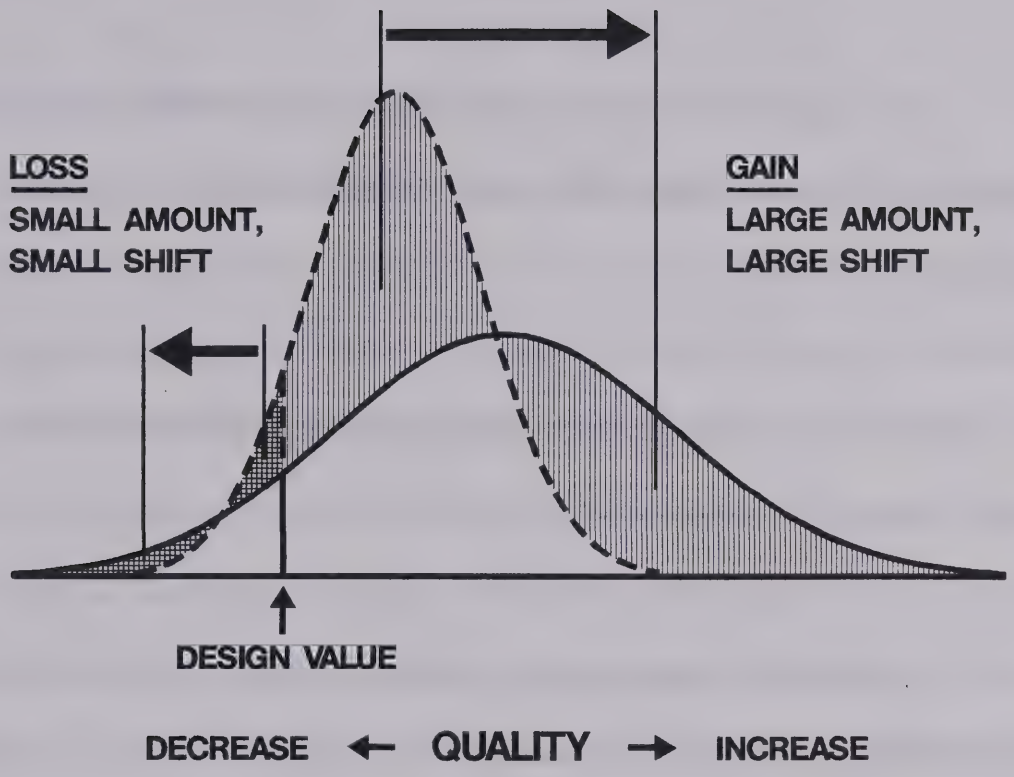


Figure 5. Illustration of case in which uniformity may not be beneficial.

for the AQL. These two distributions depict distinctly different levels of uniformity that might represent pavement strength, thickness, or other quality measure. To judge which distribution will provide better performance, observe what happens when the narrower distribution is transformed into the wider distribution:

1. The requirement that the two distributions be of the same quality (PD or PWL) causes the mean of the wider distribution to be substantially increased.

2. The quality of a small amount of material, possibly 3 or 4 percent, is slightly reduced (crosshatched areas in Figure 5).

3. The quality of a large amount of material, possibly 40 percent, is substantially improved (vertically shaded areas in Figure 5).

On the assumption that there is a gain in performance as product moves farther above a lower specification limit (thicker pavement has a greater fatigue life, etc.), the pavement represented by the wider distribution will almost certainly last longer before major repairs are needed, due largely to the increased mean value. It is very possible that it would also require less routine maintenance throughout its service life. Consequently, since the more variable distribution is expected to provide a higher level of performance, it would seem ill

advised in this situation to consider the use of a provision that would provide greater reward solely for greater uniformity. *In general, before considering the added complexity of a uniformity provision, it should be clearly established by a thorough technical analysis that such a provision is truly appropriate.*

WASTEFUL METHODS

The second diagram in Figure 1 depicts the performance of a procedure that is accurate but not precise. Procedures such as these are generally less efficient, requiring larger sample sizes for a given degree of protection (identical OC curve), or providing less protection (less discriminating OC curve) for a given sample size. Either way, they are wasteful of both staff and resources.

To provide a quantitative example showing the relative effectiveness of different types of procedures, four different acceptance plans are analyzed and compared. All four plans are designed to control percent defective or its counterpart, percent within limits. The basis for comparison is the required sample size.

Plan A is a variables plan based on the standard deviation as the measure of variability. The mean and standard deviation are computed from a random sample of N = 10 test results and the PD (or PWL) estimate is obtained from an appropriate table. To be judged acceptable, the PD estimate

must be ≤ 30 (or PWL ≥ 70). For convenience in applying procedures of this type, two new sources (6, 8) of PD/PWL estimation tables are now available.

Plan B is a variables plan based on the range as the measure of variability. The mean and range are computed from a random sample of N = 12 test results and the PD (or PWL) estimate is obtained from an appropriate table (9). To be judged acceptable, the PD estimate must be ≤ 30 (or PWL ≥ 70).

Plan C is a hybrid statistical procedure based on the average absolute deviation from the desired target value (usually the midpoint between the lower and upper specification limits). A random sample of N = 12 test results is required and, to be judged acceptable, the average absolute deviation must be no greater than 0.38 (measured as a fraction of the difference between the lower and upper specification limits for this example).

Plan D is an attributes plan, for which no statistical computations are necessary. A random sample of N = 15 test results is required, of which no more than 4 may be outside specification limits for the work to be judged acceptable.

Table 6 provides the information needed to compare the relative effectiveness of these four plans, all of which are essentially equivalent in performance as judged by their operating characteristic curves. Plan A, a conventional variables procedure based on the standard deviation, is clearly superior because it requires the smallest sampling effort with a sample size of N = 10. Plans B and C, a variables plan based on the range and a hybrid plan based on the average absolute deviation, are somewhat less efficient, both requiring a 20 percent greater sampling effort. The attributes procedure of Plan D is the least efficient, requiring a 50 percent greater sampling effort.

Table 6. Comparative performance of four acceptance plans.

	PROBABILITY OF ACCEPTANCE			
PERCENT DEFECTIVE	PLAN A (N = 10)	PLAN B (N = 12)	PLAN C (N = 12)	PLAN D (N = 15)
10	0.99	0.99	0.99	0.99
20	0.82	0.82	0.84	0.84
30	0.48	0.48	0.50	0.52
40	0.18	0.18	0.19	0.22
50	0.05	0.05	0.04	0.06
60	0.01	0.01	0.0	0.01
70	0.0	0.0	0.0	0.0

It should be noted, however, that both variables plans (Plans A and B) require that the population being sampled be at least approximately normally distributed. If this cannot confidently be assumed, it may be necessary to use an attributes plan because this requires

no distributional assumptions. Because it is not a standard statistical procedure, it is not known what distributional assumptions are required for the average absolute deviation plan.

The purpose of this comparison is to show that some statistical procedures are significantly better than others. Considerable economies can often be realized by seeking out and using the most effective methods.

IMPROVING THE VOLUNTARY CONSENSUS PROCESS

The voluntary consensus process has produced many excellent standards -- but it has also been the source of most of the ineffective procedures described in this report -- so some thought must be given toward how this process might be improved. The first logical step is to try to understand the cause of this problem. The necessary technology exists, the talent to use that technology exists, but the process that would ensure the proper use of that talent and technology has not yet been put into place.

Stephen Covey devotes an entire book (10) to the need for leaders to develop principle-centered attitudes and values to bring out the best in people and organizations. In his popular book, Quality Is Free (11), Phil Crosby notes that "If something is easy to understand and makes sense, and yet isn't done, there has to be a reason for it." He goes on to say that it usually is a failure of leadership when high standards of quality are not upheld. In his classic text, Out of the Crisis (12), Deming expresses a similar view, lamenting that today's leaders often lack a hands-on grasp of the technical issues.

Still others have made similar observations pertaining directly to the voluntary consensus process itself. The following is from Hauer, et al (13, page 20):

> The practicing engineer is relieved of the need to know (substantive knowledge) by committees that write standards, warrants, and guidelines. Then, in committee, in face of the need to act, the lack of substantive knowledge is replaced by collective wisdom. In this manner, the vital link between engineering and knowledge of fact has been cut; the artery from the heart of the profession to its brain has been severed. There is good intent without good content.

Hauer also comments on the nature of the process by which the various studies, standards, and guidelines are reviewed (14):

> The task of critical review and amalgamation remains undone. Whether it is due to misplaced civility, professional courtesy, fear of controversy, or lack of useful tradition, reviewers tend to adopt the role of an abstracting service. The result may be worse than useless. The danger is that when half-truths and myths are repeated sufficiently frequently, they acquire the ring of substantive knowledge.

Whatever the cause, it is abundantly clear that many quality assurance standards and guidelines, both in existence and under development, are replete with examples that have little in the way of solid scientific and mathematical rationale. Procedures such as these are difficult to justify, difficult to explain, difficult to defend, and certainly are not consistent with the goals of the National Quality Initiative (2).

If this situation is to improve, distinct changes must be made. The following are recommendations that, if conscientiously followed by voluntary consensus groups, would virtually guarantee the development of standards and guidelines of the highest professional integrity:

1. Select group leaders who are strongly committed to producing a quality product and who have a clear vision of how that is to be accomplished.

2. Communicate to the developers of the standard that only the correct use of the most effective methods will be acceptable. When necessary, empower and encourage them to seek the assistance of professionals who have the required technical expertise.

3. For quality assurance standards, a technical appendix describing the underlying mathematical theory is recommended. This not only forces the developers of the standard to more thoroughly understand what they are

creating, it also enables reviewers and future modifiers of the standard to better understand what the developers had in mind.

4. After the standard has been drafted, seek a thorough and independent review. This will usually involve some sort of technical analysis by a qualified professional.

The consensus process is well suited to resolve many issues, but it may not be conducive to good science. Scientific procedures must stand or fall on their technical merits. Careful scrutiny by knowledgeable individuals must be invited and welcomed, and valid criticism acted upon. Only in this manner can the process be made productive and credible.

SUMMARY AND CONCLUSIONS

The replacement value of the nation's system of roads and bridges has been estimated to be in the trillions of dollars. It was out of concern for the protection of this investment that the National Quality Initiative was created to encourage state transportation agencies to focus greater attention on all aspects of quality. However, in spite of this renewed emphasis, there remain many quality assurance procedures in use today that fall well short of conventional standards of scientific and mathematical rigor.

Many procedures are inefficient, requiring a greater sampling and testing effort than necessary, and thus are wasteful of both staff and resources. Others are ineffective, being

too weak to detect poor quality when it exists, and thus may allow seriously defective work to be accepted without the agency being aware of it. At least one procedure currently under consideration may even be counterproductive, providing greater reward for lesser quality. Poor practices such as these not only work to the detriment of the transportation profession, they may also impact adversely on the construction industry. Several examples were presented to illustrate all of these shortcomings.

There is more at issue than just a few ineffective procedures. Good science must be founded on facts and data and valid analytical techniques. If the highway profession is truly dedicated to the goals of the National Quality Initiative, it must conscientiously strive for a higher level of technical competence. The required engineering and statistical knowledge is readily available and easy to apply. What remains is for responsible leadership to provide the necessary vision, direction, and commitment.

The sources of most of the problematical examples presented in this report have been voluntary consensus standards, suggesting that there is something fundamentally wrong with the way these standards are produced. The consensus process is well suited for many things, but it may not be the best way to produce technical standards. To improve this process, conscientious leaders must see that the necessary technical expertise is made available, that a thorough review of the finished product by knowledgeable professionals is obtained, and that all valid criticism is acted upon. Only in this manner can the voluntary consensus process be made productive and credible.

REFERENCES

1. J. S. Hunter and O. J. Pendleton. On the Importance of Statistical Science in Transportation. Transportation Research Record 1340, TRB, National Research Council, Washington, DC, 1992, pages 1 - 2.

2. National Quality Initiative Long-Range Plan. Federal Highway Administration, Washington, DC, June 1994.

3. Innovation -- A Strategy for Research, Development, and Technology Transfer. American Association of State Highway and Transportation Officials, Final Draft. Washington, DC, August, 1989.

4. T. D. Larson and D. Albright. A Framework for Developing a Future Transportation Research and Development Agenda. Forum on Future Directions in Transportation R&D, TRB, National Research Council, Washington, DC, 1995, pages 159 -167.

5. S. B. Hudson. Handbook of Application of Statistical Concepts to the Highway Construction Industry, Part 1. National Technical Information Service, PB-220594, 1971.

6. R. M. Weed. Quality Assurance Software for the Personal Computer (User Manual with Computer Diskette). FHWA, Demonstration Projects Division, U. S. Department of Transportation, Washington, DC, 1996.

7. J. Windham. Implementing Deming's Fourth Point. Quality Progress, American Society for Quality Control, Milwaukee, WI, December, 1996, pages 43 - 48.

8. R. M. Weed. Managing Quality: Time for a National Policy. Report 96 - 004 - 7490, New Jersey Department of Transportation, Trenton, NJ, 1996.

9. Standard R9 - 86: Acceptance Sampling Plans for Highway Construction, Method A. AASHTO Standard Specifications for Transportation Materials and Methods of Sampling and Testing, Part I. AASHTO, Washington, DC, 1986.

10. S. R. Covey. Principle-Centered Leadership. Simon and Schuster, New York, NY, 1990.

11. P. B. Crosby. Quality Is Free. McGraw-Hill Book Company, New York, NY, 1979.

12. W. E. Deming. Out of the Crisis. Massachusetts Institute of Technology, Cambridge, MA, 1986.

13. E. Hauer, J. Lovell, and B. N. Persaud. New Directions for Learning about Safety Effectiveness. Department of Civil Engineering, University of Toronto, Report FHWA/RD - 86 - 015, Toronto, Ontario, Canada, 1986.

14. E. Hauer. A Case for Science-Based Road Safety Design and Management. Highway Safety: At the Crossroads (R. E. Stammer, Editor), American Society of Civil Engineers, New York, NY, 1988, pages 241 - 267.

NJDOT RESEARCH LIBRARY
3 4310 00011 0729